Hierarchical Bayesian Models and their Implementation in Multivariate Disease Mapping

Modelli Bayesiani gerarchici per l'analisi multivariata della distribuzione geografica del rischio di malattia

Sudipto Banerjee, Xiaoping Jin, Brad Carlin

Division of Biostatistics, School of Public Health, University of Minnesota, USA

Riassunto: La crescente disponibilità di dati georeferenziati implica anche lo sviluppo di opportuni metodi di analisi spaziale multivariata. Nel presente lavoro si propone una nuova classe di modelli generalizzati multivariati condizionatamente autoregressivi per lo studio della distribuzione geografica del rischio di malattia, quando le patologie considerate siano due o più. L'approccio sarà illustrato facendo riferimento sia all'analisi di dati puntuali, sia utilizzando dati aggregati per piccole unità amministrative (contee).

Keywords: Hierarchical Bayesian Model, Disease Mapping, Generalized Multivariate Conditional Autoregressive Model.

Summary

With the increasing popularity of Geographical Information Systems (GIS) and spatial databases, statisticians increasingly encounter multivariate spatial data. Modelling of such data require accounting for associations of more than one type and issues regarding their validity arise. From the modeler's perspective spatial data can be classified as *point referenced* or *areal*, depending upon whether it is referenced by points or areal regions or units, and their modeling presents different issues and challenges. In the fields of medicine and public health, a common application of areal data models is the study of geographical patterns of disease. When we have several measurements recorded at each spatial location (for example, information on $p \geq 2$ diseases from the same population groups or regions), we need to consider *multivariate* areal data models for handling the dependence among the multivariate components, as well as the spatial dependence between sites. In this talk, we present a flexible new class of generalized multivariate conditional autoregressive (GMCAR) models for areal data, and show how it enriches the MCAR class. Our approach differs from earlier ones in that it directly specifies the joint distribution for a multivariate Markov random field (MRF) through specification of simpler conditional and marginal models.

This in turn leads to a significant reduction in the computational burden in hierarchical spatial random effect modeling, where posterior summaries are computed using Markov Chain Monte Carlo (MCMC) methods. We compare our approach with existing MCAR models in the literature via simulation, using average mean square error (AMSE) and a convenient hierarchical model selection criterion, the Deviance Information Criterion (DIC; Spiegelhalter et al., 2002).

In the second half of this talk we attend to coregionalized areal models. While flexible modeling of multivariate point-referenced data have recently been addressed using a *linear model of coregionalization* (LMC), existing methods for multivariate areal data continue to suffer from unnecessary restrictions of less rich covariance structures (such as separability) or in undesirable dependence on the conditioning order of the variables. Here, we address these issues and propose a flexible joint modeling approach that obviates some of these drawbacks in existing multivariate areal models. We propose a Bayesian hierarchical framework for analyzing multivariate lattice data which permits modeling of correlations both between variables. Our framework encompasses a rich class of multivariate conditionally autoregressive (MCAR) models that are computationally feasible and can be compared using statistical model comparison metrics. We illustrate the strengths of our approach over existing models using simulation studies and also offer a real-data application of our proposed approach that models lung, larynx, and esophageal cancer death rates between 1990 and 2000 in Minnesota counties.

References

Spiegelhalter D.J., Best N.G., Carlin B.P., van der Linde A. (2002) Bayesian measures of model complexity and fit, *Journal of the Royal Statistical Society Series B* 64(3), 583-639.